Cutting through the complexity of the democratic backsliding problem with root cause analysis

Jack Harich (1) and Montserrat Koloffon Rosas (2)
(1) Thwink.org, Atlanta, USA, jack@thwink.org (corresponding author)
(2) Vrije Universiteit Amsterdam, Institute for Environmental Studies
Manuscript Version: November 17, 2023

Abstract: A precipitous backward slide from democracy to authoritarianism is underway. Despite decades of scholarly effort, political science lacks a comprehensive theory explaining why the problem occurs, why past solution strategies have failed, and why different future solution strategies would have a high probability of success. We argue this is because of the extreme complexity of the problem. Fortunately, problems of high complexity have long been analytically solved by the powerful tool of root cause analysis (RCA). RCA offers considerably more analytical power for complex problems than statistical analysis and experimentation, the current leading tools of political scientists. The article reviews RCA methodology, applies RCA to the backsliding problem, and offers conclusions and suggestions for further research.

Keywords: Democratic Backsliding; Authoritarianism; Root Cause Analysis; Feedback Loops; System Dynamics

Introduction

Less than thirty years after Fukuyama (1992) famously declared that since liberal democracy had proven itself superior to all other forms of government “the end of history” had arrived, democratic backsliding is strongly underway (Boese et al., 2022; Luhrmann & Lindberg, 2019; Papada et al., 2023). The problem has drawn “huge amounts of attention” from scholars (Bermeo, 2016). However, despite this effort, Waldner and Lust (2018) found that “we lack theories to explain backsliding, though we have long engaged in a perhaps interminable debate about the causes of democratic transitions, democratic breakdowns, authoritarian resilience, and democratic consolidation”. No theories were found that could persuasively explain the causes of backsliding.

Since Waldner and Lust’s review, the most comprehensive review of the literature and analysis of the problem has been the Varieties of Democracy Institute’s (V-Dem) book length study of Why Democracies Develop and Decline (Coppedge et al., 2022). Six chapters exhaustively reviewed all common theories, followed by a long chapter on their own analysis, and “is the most authoritative and encompassing empirical analysis of the causes of democratization and reversals” (italics added). The analysis applied statistics in the form of path analysis modeling.
This too is unpersuasive, because “without an experiment, a natural experiment, a discontinuity, or some other strong design, no amount of econometric or statistical modeling can make the move from correlation to causation persuasive” (Sekhon, 2009).

However, perhaps path analysis can identify major areas for further experimental investigation. Five determinates affecting downturn in polyarchy were found. Using Akoglu’s (2018) ranges for interpreting political science correlations, one determinant, lagged polyarchy, had a moderate correlation (.20). Another, anti-system movement, was negligible (-.11). The other three were none (.04, -.05, .08). According to Bellemare’s (2017) requirements for when lagged variables can be used in path analysis, polyarchy should not be used as a lagged explanatory variable because of the presence of feedback loops between the value of polyarchy and itself between years. This leaves a single explanatory relationship, anti-system movement, with a negligible correlation. Despite a wide-ranging review of many potential factors, the study was unable to find any potential causes with a moderate relationship or better, which would justify further investigation.

This lack of progress indicates current methods are unable to find problem causes, which prevents building a comprehensive theory. Why does such a large theory gap exist?

We argue the primary reason is extreme problem complexity. The democratic backsliding problem includes eight billion people, 195 countries, and thousands of state executives over the history of modern democracy. To measure problem symptoms requires over 31 million data points, over 60 indexes, and 500 indicators in one of the leading democracy indexes, that of V-Dem (Papada et al., 2023, pp. 4, 14). High problem complexity vastly exceeds the analytical capability of current political science methods.

Fortunately, problems of high complexity have long been solved by industry. At the core of their approach lies the powerful tool of root cause analysis (RCA). Since its invention by Sakichi Toyoda (whose son founded Toyota in 1937) in the early twentieth century (Ohno, 1988, p. 77), the method has gained widespread use in industry. For example, Toyota uses an RCA-based process (the Toyota Production System) to optimize the iron triangle of quality, time, and cost of 366,000 employees producing 9 million vehicles a year, each with 30,000 parts. The process has proven to be so successful it was copied (in the form of lean) and is now the global default for large manufacturing companies (Nguyen, 2018).

The basic RCA process is generic and must be wrapped in a process suitable for a particular class of problems. Examples of widely used wrapper processes are Six Sigma (Pyzdek, 2003), lean manufacturing (Womack et al., 1990), MECE issue trees (Chevallier, 2016), and fault tree analysis (Ericson, 1999). RCA processes have proven so effective that Six Sigma is used by 100% of aerospace, motor vehicle, electronics, and pharmaceutical companies in the Fortune 500 and 82% of all companies in the Fortune 100 (Marx, 2007).

This leads to our research question: Can RCA be adapted to fit the democratic backsliding problem and solve the problem at the root cause level?

Answering this question for a particular problem is so common the RCA literature provides hundreds of well-described sub-tools for construction of a wrapper process. For the backsliding problem we selected two widely used sub-tools: causal diagrams and feedback loop simulation.
modeling. These tools allowed us to identify the essential causal structure (ECS) of the backsliding problem. ECS is that portion of the complete causal structure of a system that contains the nodes, relationships (arrows), and feedback loops needed to identify the causal chains that apply to just that problem. The structure includes symptoms, intermediate causes, and root causes.³

The key findings were the main root cause and the high leverage point for resolving it. A controlled experiment was used to confirm the key findings, though this is a limited preliminary confirmation.

While we are far from a definitive explanation of the backsliding problem, we feel we have identified the backbone of the problem, in terms of the problem’s main root cause and how it can be resolved with practical solutions.

The remainder of the article reviews the theory gap, describes the RCA method, presents analysis results, and ends with conclusions and suggestions for further research.

**Review of democratic backsliding theories**

The review was conducted from an RCA point of view. Rather than an original systematic review of the literature, we rely mostly on two recent broad reviews, Waldner and Lust (2018) and V-Dem (2022).

![Causal diagrams for Waldner and Lust’s six theory families, three additional families, Fukuyama’s theory, and essential causal structure (ECS).](image)

**Figure 1.** Causal diagrams for Waldner and Lust’s six theory families, three additional families, Fukuyama’s theory, and essential causal structure (ECS).
Figure 1 summarizes review findings. Waldner and Lust’s six theory families encompass most theories, so the diagram begins with them. Additional families were added as needed. Number 10 is not a theory family but a unique theory.

None of the theories use RCA to identify the problem’s ECS. The result is ten strikingly different diagrams, even though all attempt to explain the same problem. This wide variation is what Kuhn (1996, p. 74) calls “the proliferation of theories” that occurs when a field encounters an acute anomaly, critical behaviour its present paradigm cannot explain.

The first six families are well described by Waldner and Lust.

The seventh theory family encompasses work on the role of misinformation in politics and backsliding, such as Benkler et. al.’s (2018) study of network propaganda and Lodge and Taber’s (Lodge & Taber, 2013) long running project on motivated reasoning theory. V-Dem’s 2022 Democracy Report (Boese et al., 2022, p. 7) found that “Autocratic governments increasingly use misinformation to shape domestic and international opinion in their favour.” Graves (2016, p. 6) reviews the rise of fact-checking “that seeks to revitalize the ‘truth seeking’ tradition in journalism.” All essentially say successful political deception increases backsliding. Common superficial solutions are fact-checks, articles, social media posts, news, etc. pointing out the truth.

The eighth theory family argues that backsliding occurs because the right solutions are not in place. What the right solutions are is framed in organized collections of prevention and response solutions. Catalogues of recommended policies tend to be mostly preventative, such as the Brookings Institution’s series of Democracy Playbooks (Corke, 2021). The 2021 playbook lists ten policy groups containing a total of 77 specific policies, such as “Commit to protecting and deterring undue internal (domestic) and external (international) interference in the stages of the election process.” An example of prevention and response gaps to fill is the resilience school (Merkel & Luhrmann, 2021): “Democratic resilience is the ability of a political regime to prevent or react to challenges without losing its democratic character. … There are multiple entry points to intervene…. ” Each entry point is a gap to fill with specific solutions.

The ninth theory family is described at length by Gerring (2022, pp. 55–79) in his examination of long-term “structural” causes. Long-term non-cultural forces like geography must be considered even though they cannot be changed, since they affect economic development, which in turn affects predisposition to democracy. Here “structural” does not carry the same meaning used elsewhere in this article, where “structure” refers to causal structure.

The tenth theory family is Fukuyama’s (2020) unique theory, which attempts to find the backsliding problem’s main causes. Three were found:

1. A polarizing divide based on identity politics. The left represents oppressed minorities and those who believe in quality of life for all. The right represents intolerant populist nationalism, based on ethnic superiority and belief that “our country is being taken over by a cabal of immigrants, foreign competitors, and elites who are complicit in the theft.”
2. **Appearance of the global internet and social media.** The right exploits reinforcing feedback loops that reward “conspiracy stories and fabricated information” more than the truth and encourage echo chambers of confirmation bias.

3. **The decline in authority of truthful traditional news institutions** (like newspapers, radio, and TV networks) for facts and news. The void has been filled by social media, which is much less trustworthy.

While the first cause falls into the fifth theory family, the second and third causes do not fit any theory family. Yet they offer a powerful explanation for backsliding.

We found only one recent comprehensive analysis and theory, V-Dem’s (2022) study of *Why Democracies Develop and Decline*, published in book form in 2022. The study utilized V-Dem data and “is the most authoritative and encompassing empirical analysis of the causes of democratization and reversals.” (italics added)

The study divided research on potential causes into five chapters. These correspond to the theory families of Figure 1. The chapters are geography and demographics (Geography is family 9. Demographics fits family 5.), international influences (family 6), economic factors (family 4), political institutions and democracy (family 3), and social forces and civil society with consideration of contentious politics (family 2 and 5). The final chapter synthesized previous chapter findings into a comprehensive theory using path analysis modeling and attempted to find causes by statistical means. As described in the Introduction, the study was unable to find problem causes.

The main point of Figure 1 is that while all the theory families have useful concepts, none attempt to explicitly find the problem’s ECS. Instead, all are collections of related factors and plausible intermediate (proximate) causes loosely connected by problem stories. Their conclusions are intuitively derived, causing ten completely different theories and diagrams, despite the fact that all attempt to explain the same problem. All lack a viable path to solution. Because the root causes were not found, the high leverage points for solution application are unknown.

**Method and definitions**

The RCA paradigm rests on several foundational definitions. Drawing from a diversity of sources, e. g. (Andersen & Fagerhaug, 2006; George et al., 2005; Ishikawa, 1986; Okes, 2019; Pyzdek, 2003; Tague, 2005), *root cause analysis* (RCA) is the systematic practice of finding, resolving, and preventing recurrence of the root causes of causal problems. A *root cause* is the deepest cause in a causal chain (or the most basic cause in a feedback loop structure for more complex problems) that can be resolved by changing something in the cause, such as stopping it, increasing it, or fixing it. *Resolved* means the problem will probably not recur due to that root cause.

A *causal problem* occurs when problem symptoms have causes, such as illness or a car that won’t start. Examples of non-causal problems are information search problems, math problems,
scientific discovery problems, and puzzle solving. The golden rule of RCA is *All causal problems arise from their root causes.*

For difficult large-scale social problems, the above definition of root cause is too vague. A rigorous strong definition is required: A *root cause* is that portion of a system’s feedback loop structure that, using the checklist below, explains why the system’s structure produces a problem’s symptoms. The checklist allows numerous unproductive root causes (particularly intermediate causes posing as root causes) to be quickly eliminated. The *five requirements of a root cause* are:

1. It is clearly a (or the) major cause of the symptoms.
2. It has no worthwhile deeper cause. This halts the asking of “WHY did this occur?” at an appropriate point.
3. It can be resolved, by pushing on its high leverage point(s) to initiate the desired mode change (described later) in complex problems, or to merely change the node with the root cause in simple problems. (Mode change versus node change)
4. Its resolution will not create other equal or bigger problems. Side effects must be considered.
5. There is no better root cause. All alternatives have been considered to the point of diminishing returns.

The first three requirements were specified by (Harich, 2010). In the spirit of continuous process improvement, two more have been added. An example of a tool for satisfying requirement five is MECE Issue Trees (Chevallier, 2016), where each causal chain layer is searched using mutually exclusive collectively exhaustive causes.

An *intermediate cause* is any node on the causal structure between symptoms and root causes. There are two signs a node is an intermediate rather than a root cause:

1. Using a causal diagram or simulation model, it’s easy to see how there is a deeper cause. Or if the diagram or model is under construction, it’s easy to list possible deeper causes.
2. When solution elements attempt to resolve a cause the solutions unexpectedly fail, despite ingenious solution designs and prolonged effort. This behaviour indicates deeper forces are at play. Those deeper forces must be a deeper intermediate cause or a root cause.

Leverage points are solution strategies. A *low leverage point* is a node offering a practical place for *superficial solutions* to “push” in order to resolve an intermediate cause. A *high leverage point* offers a place for *fundamental solutions* to push to resolve a root cause. A high leverage point is connected to a root cause in such a manner that pushing on the high leverage point reduces the root cause force to an acceptable level or eliminates it altogether. This resolves the old root cause forces and creates new root cause forces.
RCA in its most basic form consists of the *Five Whys* (Liker, 2004, pp. 252–256). The method asks “WHY does this occur?” until the most basic cause, the root cause(s), is found. Solutions are then designed, tested, and applied to resolve the root cause.

RCA arose from the Toyota Production System, whose foundation is the Five Whys. To explain the Five Whys, Ohno (1988, p. 17) gives this widely cited example:

…suppose a machine stopped functioning. [problem symptoms]

1. Why did the machine stop?
   - There was an overload and the fuse blew. [intermediate cause 1]
2. Why was there an overload?
   - The bearing was not sufficiently lubricated. [intermediate cause 2]
3. Why was it not lubricated sufficiently?
   - The lubrication pump was not pumping sufficiently. [intermediate cause 3]
4. Why was it not pumping sufficiently?
   - The shaft of the pump was worn and rattling. [intermediate cause 4]
5. Why was the shaft worn out?
   - There was no strainer attached and metal scrap got in. [root cause]

Without the strict Five Whys procedure, problem solvers tend to stop long before the root cause is found and apply a superficial solution, such as replacing the fuse or the pump. RCA forces problem solvers to *think analytically in a systematic productive manner* by relentlessly asking WHY until the root cause is found, which opens the door to fundamental solutions to resolve the root cause.
Using standard RCA terminology, the Five Whys example may be diagrammed to illustrate how ECS works (Figure 2). The diagram shows the four main forces driving the behaviour of difficult causal problems: force S, F, R, and New R. On the diagram, S < R signifies that superficial solutions cannot resolve intermediate causes, since force S is always less than force R. F > R signifies that properly designed and tested fundamental solutions can resolve root causes, since force F exceeds force R. The desired system mode change occurs when the root cause is resolved and New R appears. Note how clearly and completely the diagram describes the problem’s ECS in an engineering-like manner.

Figure 2. Essential causal structure (ECS) of the machine stopped problem.
All reliable methods for solving causal problems use some form of RCA, whether RCA terminology is used or not. For example, doctors use RCA but not its terms. A patient presents with symptoms of chronic knee pain. If examination and tests show the root cause is severe knee osteoarthritis, then the high leverage point is knee surgery. Fundamental solutions are different surgery types. But suppose the root cause was unknown or the patient tries to self-diagnose and self-treat. A low leverage point might be anti-inflammatory drugs to reduce the intermediate cause of inflammation. Superficial solutions would be different drugs like aspirin, ibuprofen, or steroids. None of the superficial solutions could ever permanently resolve the intermediate cause because of the deeper root cause. Once the root cause was resolved, if chronic pain had been causing a downward spiral of severe depression, drug addiction, or other dysfunction, resolution would lead to a dramatic system mode change, where quality of life soared from low to high due to changing to a virtuous upward spiral (a reinforcing feedback loop with growth instead of decline).

RCA is similar to causal process tracing (CPT) but is an older more mature method for identifying causal mechanisms, and is more capable of efficiently finding ECS. In particular, CPT lacks the all-important concept of using the Five Whys to go past intermediate causes to root causes. As described at length by Beach and Brun (2012, p. 16), there is only tracing “backward from Y to uncover a plausible X.” Without strong differentiation between intermediate and root causes, which is the heart of the RCA paradigm, attempts at causal inference on difficult social problems tend to fall into the superficial solutions trap. This occurs when due to extreme problem complexity, people assume intermediate causes are root causes. It is a common trap, as Forrester describes: (italics added)

*The intuitively obvious ‘solutions’ to social problems are apt to fall into one of several traps set by the character of complex systems. ...people are often led to intervene at points in a system where little leverage exists and where effort and money have but slight effect.*

...social systems are inherently insensitive to most policy changes that people select in an effort to alter behavior. In fact, a social system draws attention to the very points at which an attempt to intervene will fail. Human experience, which has been developed from contact with simple systems, leads us to look close to the symptoms of trouble for a cause. But when we look, we are misled because the social system presents us with an apparent cause that is plausible according to the lessons we have learned from simple systems, although this apparent cause is usually a coincident occurrence that, like the trouble symptom itself, is being produced by the feedback loop dynamics of a larger system.

Forrester’s “apparent cause” is what RCA calls the intermediate cause. “Little leverage exists” if problem solvers assume the apparent cause is the root cause, because that leads to pushing on low leverage points.
Figure 3. Completed ECS diagram of the democratic backsliding problem. A feedback loop simulation model (Figure 4) contains the detailed ECS for the fundamental layer. By coincidence, Figures 2 and 3 both have four intermediate causes.

Analysis results are summarized in Figure 3. Where theory families 1 to 7 in Figure 1 fit into the analysis is shown at the top of each of the four superficial layers. Family 8 fits in all four layers. Family 9 fits nowhere since it has little impact. Family 10 contains Fukuyama’s three main causes. The first fits into family 5. The other two relate to the fundamental layer.
The diagram was developed by asking a methodical series of WHY questions:

1. **The first WHY question:** After summarizing problem symptoms as *backsliding from democracy to authoritarianism*, we asked: WHY do those symptoms occur? Because of *backsliding decisions made by politicians*. Bermeo (2016) describes how “Executive aggrandizement… occurs when elected executives weaken checks on executive power one by one, undertaking a series of institutional changes that hamper the power of opposition forces to challenge executive preferences.” These decisions are made due to various *politician behavior factors*. To change that, various solutions designed to *promote moderate preferences and commitment to democracy* are used. These superficial solutions usually fail, because they do nothing to change why politicians making backsliding decisions are in power.

2. **The second WHY question:** Next, we asked: WHY do backsliding decisions made by politicians occur? Svolik (2019) reports that 197 democratic backslides occurred from 1973 to 2018. Of these, 46 were military coups and 88 were executive takeovers via election, with takeovers averaging about 80% of all backslides after the end of the Cold War in 1991. In an examination of *How Democracies Die*, Levitsky and Ziblatt (2018) summarize this change: “Democratic backsliding today begins at the ballot box.” Thus, the main second intermediate cause of backsliding is *election of politicians not working for the democratic common good*. Too many citizens are voting instead for politicians working for the uncommon good of powerful special interests, such as authoritarians.

   To solve that intermediate cause, the solution strategy used now is the same one used before to hasten the spread of democracy before the backslide began: *more of the truth via promotion of proof of superiority of democracy over autocracy*. Solution elements are books, articles, projects, etc., such as the book *The Democracy Advantage* (Halperin et al., 2010). This is a form of “more of the truth” and is part of classic activism (Harich, 2010), the traditional general-purpose process used to solve public interest problems. More of the truth is a low leverage point.

   The solution no longer works because since about 2000, authoritarian state capacity (government effectiveness, regulatory quality, rule of law, and control of corruption) has improved so much worldwide, notably in China, that there is no longer proof democracy is superior. “Of the twenty fastest growing countries of the past two decades, fifteen have been autocratic regimes. Of the fifteen wealthiest economies in the world today by per capita income, almost two-thirds are nondemocracies” (Foa, 2018).

3. **The third WHY question:** Next, we asked: WHY does *election of politicians not working for the common good* occur? Mainly because of *pernicious polarization of voters*. Somer, McCoy, and Luke (2021) examined this topic and found that pernicious polarization is “the division of society into mutually distrustful Us versus Them camps in which political identity becomes a social identity. [This] fosters autocratization by incentivizing citizens and political actors alike to endorse nondemocratic action” and elect nondemocratic politicians.
Svolik (2019) asked the same WHY question we did: “Why do voters support politicians who undermine democracy?” and concluded that “In sharply polarized electorates, even voters who value democracy will be willing to sacrifice fair democratic competition for the sake of electing politicians who champion their interests.” A controlled experiment found that voters “are indeed willing to trade off democratic principles for partisan interests.”

How can polarization be overcome? Somer, McCoy, and Luke (2021) “argue that democratic resilience must include capacities to prevent or reverse pernicious polarization that erodes democracies and strengthens autocrats,” which is the low leverage point in the ECS diagram. Proposed solutions for pushing on that leverage point are various goals and strategies that actors opposing polarization can use. Somer, McCoy, and Luke “offer a framework with two broad opposition goals and four different strategies that oppositions may adopt to respond to polarizing incumbents with democracy-eroding behaviors.” These consist of proactive counter-polarization strategies, transformative repolarization, and active depolarization strategies.

Will solutions like these succeed? We expect they will continue to have only a modest effect at best, because they do nothing to address why polarization appears. They only attempt to reduce polarization after it appears.

4. The fourth WHY question: Next, we asked: WHY does pernicious polarization of voters occur? Mainly because of successful political deception. Other techniques are far less effective. As explained later in the section on Feedback loop simulation model, deception is the main technique used to convince an electorate majority to vote against their own best interests.

All authoritarians depend on deceptively provoking a wide range of false beliefs and related emotions, especially fear. McCarthy (2019) describes how this involves “bellicose rhetoric,” false promises of forceful solutions to complex long-term problems, populist and racist appeals, fanning the flames of fear and hate against false internal and external enemies, and more. Walker (2016) documents how authoritarian regimes employ increasingly sophisticated forms of deception, such as control of the media to spread state propaganda, simulated democratic institutions like fake or weak civil society organizations and political parties, censorship of the truth, etc.

Without an RCA mindset, intermediate causes are routinely assumed to be root causes, as has occurred here for decades. If the “cause” is successful political deception, then the solution is intuitively obvious: some form of more of the truth via misinformation correction, a low leverage point. The reasoning is that if people believe statements that are not true, then that can be corrected by providing citizens with corrected versions of deceptive statements pointing out the truth. Superficial solutions doing this are fact-checks, articles, social media posts, news, etc. pointing out the truth.

Ample research shows misinformation correction works poorly (Persily & Tucker, 2020, pp. 163–198), so poorly that we are now living in the post-truth age of politics, “in which lies and distortions carry as much weight as facts” (Puddington, 2017). “The post-truth politician manufactures his or her own facts. The post-truth politician asserts whatever they believe to be in their own interest and they continue to press those same claims, regardless of the evidence
amassed against them” (Lockie, 2017). Recent examples are the “policy deception” behind the disastrous Brexit vote (Baines et al., 2020), Donald Trump’s 30,573 false or misleading claims while in office (Kessler et al., 2021), and Putin’s lies about Russia’s invasion of Ukraine (Myers & Thompson, 2022).

While intuitively it should work, in practice misinformation correction has little effect for two main reasons: (1) Motivated reasoning research has found that partisan beliefs are highly resistant to change once formed. “The motivational component of political misinformation implies that the prospects for correcting false beliefs are dim” (Jerit & Zhao, 2020). (2) Misinformation correction has never solved the problem of how to get corrections to the people that need them. “We almost never observe respondents reading a fact-check of a specific claim in a fake news article that they read” (Guess & Nyhan, 2018). This explains why misinformation correction is a low leverage point.

**Feedback loop simulation model**

5. **The fifth WHY question**: Finally, we asked: WHY does successful political deception occur? Answering this was so difficult it required construction of a feedback loop simulation model (Figure 4). We elected to build a small insight model rather than a much larger more complex calibrated model to greatly speed development and simplify model structure. Either is capable of finding all ECS, as long as RCA drives model construction. Small insight models:

“…are unique in their ability to capture important and often counterintuitive insights relating behavior to the feedback structure of the system without sacrificing the ability for policymakers to easily understand and communicate those insights. …for many public policy problems a small model is sufficient to explain problem behavior and build [understanding] regarding appropriate policy responses” (Ghaffarzadegan et al., 2011).

Axelrod (1997), one of the early pioneers of simulation in the social sciences, makes a further distinction regarding choice of model characteristics. **Model accuracy** is how closely a model mimics the real world, in terms of its structure and behavior. If high realism and accurate prediction is needed, such as in an economic model predicting interest rates or a simulator for training a supertanker crew, then: (italics added)

…*accuracy* is important and *simplicity* of the model is not. But if the goal is to deepen our understanding of some *fundamental* process, then *simplicity* of the assumptions is important and *realistic representation* of all the details of a particular setting is not.

The strategic goal of RCA is understanding a problem’s fundamental layer behavior, particularly where the high leverage points are and whether pushing on them will alter system behavior enough to solve the problem. Accurate prediction of behavior is not required unless one is optimizing rather than changing behavior so disruptively that a mode change occurs. Model simplicity is thus preferred. The simpler model structure is, the faster process step iterations will
The more easily everyone can grasp analysis results, and the more easily model structure can be validated by comparing it to the real world.

Figure 4. Causal loop diagram of the democratic backsliding problem’s fundamental layer ECS. This is a high-level diagram of the system dynamics model. Here “degenerate” is not pejorative, but signifies a person has fallen from the norm of rationality. They have degenerated into supporters of a false ideology. Solid lines are direct relationships. Dashed lines are inverse relationships. Dotted lines are constants. Gray text is comments. See the appendix for description of the simulation model and scenario runs.

The Dueling Loops of the Political Powerplace model captures the essence of the left-right political spectrum, consequential because “global politics is first and foremost a debate between the left and the right. ... The left-right dichotomy occupies a special place, as the most enduring, universal, and encompassing of all political strategies” (Noel & Therueb, 2008). The backbone of the model is the two opposing feedback loops dueling for the same Uncommitted Supporters. Race to the Bottom politicians (the right) use deception to gain supporters, while Race to the Top politicians (the left) use the truth.
The model uses Dawkins’ (1976, pp. 189–201) concept of memes. A meme is copied information capable of affecting behavior, such as a fact or an opinion. In the model a meme is a statement that is true or false.

Truth literacy is the ability to tell truth from deception, i.e., to be able to “read” the truth. The analysis uses three variables we developed to measure components of political truth literacy. All range from zero to 100%:

- **LTQ** (*logical truth quotient*) is the ability to logically tell if a political claim is true or false.
- **AAQ** (*appropriate action quotient*) is the ability to take appropriate action (especially voting correctly), given the perceived truth (using LTQ) of a political claim.
- **DTQ** (*democratic truth quotient*, aka political truth literacy) is the ability to take correct democratic system action given a political claim.

A person’s DTQ uses the two-step process of (1) determine the truth (using LTQ) and then (2) take action given that perceived truth (using AAQ). For example, “I can see that statement is false because it uses the cherry-picking fallacy.” And then “Now I need to vote against that politician because they cannot be trusted to tell me the truth.” Because of the two-step process, the three variables are related by DTQ = LTQ x AAQ, though DTQ is not in the model. The single high leverage point in the completed ECS diagram (Figure 3), raise political truth literacy from low to high, becomes two high leverage points in the simulation model: LTQ and AAQ.

In an ideal world no one would lie. But real-world politics is full of lies to gain more supporters. Deceptive (degenerate) politicians increase the attractive power of political statements by lying, represented by *false meme size* of greater than one. This gives The Race to the Bottom an advantage, but only if LTQ and AAQ are low. If they are high, most lies are detected and The Race to the Top has the advantage.

The Race to the Bottom feedback loop (the right) represents powerful special interests pursuing their own narrow self-interest goals, such as the rich, managers of large for-profit corporations, authoritarians, and elite ruling groups of many kinds, e.g., the ruling class. All are a small percentage of the electorate.

In a democracy, the main ways a minority can persuade a majority to vote for them are by force, threats, rigged elections, voter suppression, favoritism, bribes, or deception. Force, threats, and rigged elections are illegal. Voter suppression is mostly illegal. Favoritism doesn’t work on large populations, since there are not enough favors (like jobs or contracts) to dole out. Bribes are inefficient, as even the rich lack the resources to bribe millions of voters. This leaves deception as the main preferred strategy and explains why deception is so common in right-wing politics and why *successful political deception* is intermediate cause 4 in Figure 3. Jeremy Bentham, the father of utilitarianism, reached the same conclusion in 1824: (Larrabee, 1925, p. xxi)
…it is impossible by fair reasoning ...to justify the sacrifice of the interests of the many to the interests of the few.... It follows that for effecting this purpose they must have recourse to every kind of fallacy, and address themselves, when occasion requires it, to the passions, the prejudices, and the ignorance of mankind.

The Race to the Top feedback loop (the left) represents those seeking to cooperate in optimizing the long-term common good of all. Politicians appealing to the left use a strategy of the truth about how they can achieve that goal.

The low leverage point of more of the truth via misinformation correction on the ECS diagram (Figure 3) is an attempt to increase the attractive power of true memes on the simulation model (Figure 4). While intuitively this should work, it has little effect as seen in the continued failure of superficial solutions. This occurs due to the unresolved root cause.

Identification of the main root cause

The key model insight is that the size (and hence the attractive power) of a lie (false memes) can be inflated, while the size of the truth (true memes) cannot. From a mathematical perspective, the size of a falsehood can be inflated by saying that $2 + 2 = 5$, or 7, or even 27, but the size of the truth is always 1. It can never be inflated by saying anything more than $2 + 2 = 4$. Inflation is used to create fear when there is nothing to fear, doubt when there is nothing to doubt, the false promise of I can do so-and-so for you when I really cannot, a large flaw in one’s opponent when there is only a small flaw or no flaw, etc. This insight leads to identification of the main root cause of backsliding: the inherent advantage of the Race to the Bottom, represented on the model by undetected false memes. The inherent advantage exists because the opposing loop, the Race to the Top, has no corresponding node because there are no inflated true memes to detect. For simplicity we usually say the main root cause is low political truth literacy.

If this is the main root cause, one would expect to see universal reliance on political deception by right wing governments and politicians. Evidence confirming this is considerable.

Observational confirmation

China and Russia are infamous for dependence on state-sponsored propaganda, both internal and external (Diamond, 2020). In the US a well-financed “Right-Wing Propaganda Machine” has dominated political debate for decades (Conason, 2004). These are isolated cases, however. A systematic examination of the evidence is required.

Using the V-Party dataset (which covers 1,943 political parties across 1,759 elections in 169 countries from 1970 to 2019), Luhrmann et. al. (2021) identified four key characteristics of anti-pluralism. All require political deception to implement:

1. Unwillingness to commit to the democratic process as legal means for gaining power.
   Merloe (2016) found “The spread of disinformation… is essential to authoritarian attempts to control the electoral narrative” in order to allow rigged election results to
be accepted. For example, The Big Lie of Trump that the US 2020 election was stolen uses deception to foster the idea that bypassing the democratic process is okay, since the election process is rigged against Trump and Republicans.

2. **Denial of the legitimacy of dissenting parties and opponents.** Luhrmann et. al. (2021) describe the standard approach: “delegitimize, severely personally attack, or demonize [or dehumanize] their opponents.” This is accomplished with lies about why the opposition is not legitimate, ad hominem attacks, and lies to demonize and dehumanize.

3. **Toleration, encouragement, or endorsement of the use of violence against political opponents.** Deception is required to justify violence against people who are not dire threats. Doing this takes two steps. The first is accomplished with the second characteristic by deceptively portraying the opposition as illegitimate and inhuman. The second step deceptively justifies violence as the best means to eliminate or intimidate that opposition. Examples are Mussolini’s blackshirts, Hitler’s brownshirts, and Trump’s incitement of the Oath Keepers, Proud Boys, and other groups and individuals in the January 6, 2021 US Capitol insurrection. In every case, deceptive rhetoric is used to justify and motivate. Mussolini championed the “virtue” and “supreme morality” of violence against socialists and other demonized opponents, carried out by the blackshirts (Ebner, 2010). In 1932 Germany, the Nazi press justified the infamous murder of Pietrzuch, “a Polish rogue and sub-human,” by a gang of brownshirts “as an act of lynching—a practice [that was] the only possible corrective to an unnatural law” (Siemens, 2017).

4. **Support for curtailing the civil liberties of minority groups.** This is accomplished by falsely painting a minority group as a dire threat, which justifies curtailing their civil liberties. Common scapegoats are religious groups, racial minorities, immigrants, and LGBTQ people. Examples are Hitler’s extermination of the Jews, Russia’s homophobic laws (Kurtanidze, 2021), and the rise of far-right racist populism in Europe, the US, and Australia (Vieten & Poynting, 2016).

Evidence of the two Dueling Loops itself exists. Benkler et. al. (2018) in an effort to understand the effect of propaganda on politics, analyzed four million messages in the US using their Media Cloud platform. America’s political spectrum has evolved into two opposing feedback loops, a right-wing “propaganda feedback loop” where politicians “compete on identity confirmation” regardless of the truth, versus a centrist/left-wing “reality-check” loop that follows “institutionalized truth-seeking norms” where politicians “compete on truth quality and the scoop”. The propaganda and reality-check loops correspond exactly to the Race to the Bottom and Top loops.

Freelon et. al. (2020) found that “in the US and throughout the industrialized West… available evidence suggests that the right has invested far more than the left in disinformation and conspiracy theories as core components of its activist repertoire…. Also in the US, while both
parties use forms of deception like anti-elitism, anti-immigration, and demonization of political opponents, the data show Republicans do so much more frequently than Democrats (Luhrmann et al., 2020).

Fukuyama’s three main causes of backsliding, listed earlier, support the Dueling Loops structure: (1) The “polarizing divide” is the two opposing loops. (2) The amplification of certain false memes by social media has increased the attractiveness of false memes. (3) The decline of traditional media (which filtered out most lies and provided citizens with the truth) and the rise of social media (which does the opposite) has caused the number of false memes to be amplified. Citizens are exposed to many more false memes.

The effect of social media false meme amplification has become large and continues to grow (Fisher, 2022). This amplification gives the Race to the Bottom a further advantage because of the unresolved root cause of low political truth literacy. Once the root cause is resolved, amplification no longer works. See simulation run 13 in the appendix for further examination of amplification.

**Experimental confirmation: The Truth Literacy Training study**

A collection of six solution elements was designed to push on the high leverage point. The one with the highest probable impact for the least amount of effort appears to be Truth Literacy Training. This trains citizens on how to tell political truth from deception and then use that knowledge to make important political decisions correctly, especially voting.

To experimentally test the hypothesis that the main root cause exists and can be resolved, we performed the Truth Literacy Training study, to be fully described in a future article after this article has established the root cause. Here we offer a brief summary.

Using a Prolific online panel with United States participants, subjects were randomly assigned to treatment groups. Demographics were age range 22 to 51, average age 31, 49% male. Educational levels were 34% high school, 55% college degree, 10% PhD. The trained group (33 subjects) received Truth Literacy Training on how to spot political deception by drilling on a catalogue of common fallacies, how to use of the Personal Truth Test, and how to take appropriate action. The control group (30 subjects) was exposed to a neutral topic.

Results were positive. Average DTQ for the control group was very low, 2%, with a 95% confidence range of zero to 10%. This offers initial confirmation the root cause of low political truth literacy exists. Average DTQ for the trained group was 67%, a 65-point rise. Training averaged one hour. This suggests that Truth Literacy Training is capable of pushing on the high leverage point of raise political truth literacy from low to high successfully, though replication, larger sample size, and much further research is required. We view these results as tentative rather than conclusive.

A follow up questionnaire 26 days later showed DTQ for the trained group dropped slightly, from 67% to 60%. This rose to 70% with 30 minutes of refresh training, indicating that regular refresh training of some type can work and will be required. Or it may be that like reading and
writing literacy, once political truth literacy matures, becomes the reasoning default and is ex-
ercised often enough, little decline will occur.

Conclusions

The article asked: *Can RCA be adapted to fit the democratic backsliding problem and solve the problem at the root cause level?* Analysis results suggest it can. We draw several conclusions with significant implications:

1. The completed ECS diagram of Figure 3 offers a comprehensive theory of the democratic backsliding problem by explaining the problem in terms of the four main causal forces involved:

   **Force S. Why past solutions failed** (force $S < R$). Due to lack of appropriate analytical methods, the fundamental layer of the problem was hidden by complexity. This caused problem solvers to be intuitively attracted to pushing on low leverage points with superficial solutions.

   **Force R. Why the problem occurs** (force $R$ is unresolved). Susceptibility to backsliding occurs because of the unresolved main root cause of low political truth literacy. This hypothesis is well supported by the literature and on a preliminary basis by experiment.

   **Force F. Why fundamental solutions can be expected to succeed** (force $F > R$). The long history of RCA has established the fact that all causal problems arise from their root causes, and that root causes can be routinely resolved with the appropriate mature process. If a root cause is identified using a rigorous form of RCA and the root cause hypothesis is tested by measurement and experimental application of solution elements, there is a high probability that the full-scale solution will work, though solution element adjustment is usually required.

   **Force New R. Why the mode change, and hence the solution, will be relatively permanent** (force $F$ causes force $R$ to transition to force New $R$). Simulation model scenarios show that the system undergoes a striking systemic mode change once the root cause is resolved and political truth literacy changes from low to high. The tendency of politicians to tell lies does not just fall to a low level. *It disappears altogether, because now the winning strategy for politicians is telling the truth.* Those who do not tell the truth die out. See simulation run 10 in the appendix, which demonstrates this behavior. The effect is analogous to the eradication of smallpox, which itself succeeded due to RCA, although RCA terminology was never employed (Bazin, 2000).

2. From the viewpoint of the method used, the above four aspects can be considered the *four requirements* for a convincing comprehensive theory. While this may seem like an excessively high bar, we see no other minimum set of requirements capable of
specifying the information needed to solve the backsliding problem, due to its extreme complexity and long history of repeated solution failure.

3. No other theory we are aware of meets any of the four requirements.

4. The comprehensive theory suggests the democratic backsliding problem is realistically solvable, once high leverage point solution elements for resolving the main root cause are developed, tested, and applied, and solutions are iteratively improved.

Suggestions for further research

These conclusions suggest further areas of research.

One area is a calibrated simulation model of the fundamental layer of the problem. A model is calibrated by measuring factors in the real world, using those values in the model, and running the model to compare its dynamic behavior to that in the real world. Model structure is then improved as needed to narrow the gap between model and real-world behavior to an acceptable level. The great advantage of calibrated versus estimated simulation models (such the estimated model used in this article) is a calibrated model offers much stronger validation and much closer prediction. The latter is useful when weighing alternative solution strategies. Both allow efficiently exploiting solution behavior results to iteratively improve the model. At this point the analysis becomes self-validating and self-improving. As explained below, such a calibrated model would signal emergence of an effective methodological paradigm that marks a portion of political science as mature.

Every scientific field confronts difficult problems. Every mature field has improved its central problem-solving methods until they could routinely produce reliable calibrated simulation models, broadly defined as system models where causal structure is identified and quantified, such that model behavior reproduces or predicts real world behavior to the desired level of accuracy. Once calibrated models can be easily produced, a field can routinely solve its central problems. Physics famously pioneered this approach with Newton’s Three Laws of Motion and the Law of Universal Gravity, which allowed quantified models of motion problems over time to be easily created. Similar approaches were taken with chemical reaction models, evolutionary biology’s models of species change, epidemiology’s contagious disease models, weather prediction models, economic models, climate change models, industrial process models, financial models, and countless other fields and applications. Calibrated simulation models have become the gold standard that marks a science as mature.

Once political science creates its first calibrated model of a solved problem central to the field, it can stand alongside economics as a semi-hard science. Based on analysis results, the long history of mature RCA processes solving industry’s toughest problems one by one, and the “huge amounts of attention” from scholars (Bermeo 2016) on the backsliding problem, we feel this goal is imminently close to achievement.

A second area would be further work on the truth literacy training experiment. Can it be replicated? What sorts of training would be required to inoculate the average citizen against
political deception in a particular country, as tested by experimentation? Is it possible to target just those citizens who would benefit the most from higher political truth literacy, such as swing voters and those so young they have not yet been indoctrinated by a deception-based political ideology? What training works best against particular deception strategies, like authoritarian populism, political cults,\(^5\) false common enemies (such as immigrants or non-whites), and pushing the fear hot button?

Finally, we foresee significant potential in further work on a collection of solution elements designed to optimize pushing on the high leverage point. To illustrate the strategy six solution elements were designed: Freedom from Falsehood, No Competitive Servant Secrets, Politician Truth Ratings, Politician Corruption Ratings, Sustainability Quality of Life Index, and Truth Literacy Training. The elements work together to dynamically enhance the speed and effectiveness of pushing on the high leverage point, using carefully engineered feedback loops such as Lifting the Blanket of Deception, The Winning Strategy Is Tell the Truth, The Drive for Politician Rating Excellence, and Truth Literacy Promotion. The feedback loops cause the system to inherently strive to reach two new explicit goals: the sub-goal of High Political Truth Literacy and the top system goal of Long-Term Optimization of the Common Good. The latter summarizes the goal of modern democracy.\(^7\) If this research vision and the related required mode change can be implemented, then that goal is now the system’s top explicit goal and hence is likely to be achieved. Such advanced solution strategies are impossible without deep knowledge of system ECS and feedback loop simulation modeling.

For how political and social scientists can begin applying RCA using ECS diagrams, please see the appendix. Given analysis results, political scientists have an unfair head start!

**Acknowledgements**

(To be completed later.)

**Appendix**

The appendix describes the system dynamics simulation model and scenario runs, and further work: (1) If political truth literacy has always been low, why didn’t backsliding begin long ago? (2) Is a single high leverage point too simplistic for such a difficult problem? (3) How political and social scientists can apply RCA using ECS diagrams. (4) The transference research program. The Vensim simulation model is in a separate file.

**Notes**

\(^{1}\) A correlation of .7 and up is very strong, .4 to .69 is strong, .20 to .29 is moderate, .10 to .20 is negligible, and zero to .10 is none

This definition of ECS is similar to that used in the social sciences (Morgan, 2013, p. 279): “For a given causal process, we do not want an account that includes all details of the process, but rather one that captures only the crucial causal elements while leaving others out.”

Here “difficult” means problem complexity is so high, and the problem has resisted repeated solution for so long, that a workable solution requires application of an appropriate form of RCA capable of identifying the four forces shown on the ECS diagrams of Figures 2 and 3. The method must fit the problem.

See https://www.prolific.co.

Political cult examples are the Nazi Party, Kim Jon-un’s North Korea, Vladimir Putin’s Russia, Xi Jinping’s remoulding of Chinese communist party, and “The Cult of Trump” as described by Hassan (2019).

That the goal of modern democracy is to optimize the long-term common good may be seen, for example, in Dahl’s (2017, p. 45) list of the ten desirable consequence democracy produces. Number eight, “political equality,” is the “common” in the common good. The other nine consequences increase quality of life, the “good” in the common good. Dahl’s discussion of the ten consequences implies their optimization, via phrases such as “a broader range of freedom than any feasible alternative” (p50), “a maximum opportunity for persons to exercise the freedom of self-determination” (p53), and “democracy fosters human development more fully than any feasible alternative” (p55). Or as Reich (2018, p. 13) explains, the goal of the “common good” is enshrined in the U.S. constitution, whose purpose was to “promote the general welfare” of “we the people.” The U.S. constitution established the world’s first modern democracy and served as the model for all others. The purpose of democratic constitutions is to define the mechanisms for implementing democracy indefinitely, which implies the “long-term” aspect of the goal.
References


